



Superoxide dismutase 1 is positively selected to minimize protein aggregation in great apes

Dasmeh, Pouria; Kepp, Kasper Planeta

Published in:
Cellular and Molecular Life Sciences

Link to article, DOI:
[10.1007/s00018-017-2519-8](https://doi.org/10.1007/s00018-017-2519-8)

Publication date:
2017

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Dasmeh, P., & Kepp, K. P. (2017). Superoxide dismutase 1 is positively selected to minimize protein aggregation in great apes. *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-017-2519-8>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Superoxide dismutase 1 is positively selected to minimize protein aggregation in great apes

Pouria Dasmeh^{1†}, and Kasper P. Kepp^{*2}

¹ *Harvard University, Department of Chemistry and Chemical Biology, Cambridge, MA, USA*

² *Technical University of Denmark, DTU Chemistry, DK-2800 Kongens Lyngby, Denmark.*

[†] *Current address:* Department of Biochemistry and Cedergren Center for Bioinformatics and Genomics, Faculty of Medicine, University of Montreal, 2900 Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada.

^{*} Corresponding Author e-mail: kpj@kemi.dtu.dk Phone: +045 45 25 24 09

Abstract

Positive (adaptive) selection has recently been implied in human superoxide dismutase 1 (SOD1), a highly abundant antioxidant protein with energy signaling and anti-aging functions, one of very few examples of direct selection on a human protein product (exon); the molecular drivers of this selection are unknown. We mapped 30 extant SOD1 sequences to the recently established mammalian species tree and inferred ancestors, key substitutions, and signatures of selection during the protein's evolution. We detected elevated substitution rates leading to great apes (Hominidae) at ~1 per 2 million years, significantly higher than in other primates and rodents although these paradoxically generally evolve much faster. The high evolutionary rate was partly due to relaxation of some selection pressures and partly to distinct positive selection of SOD1 in great apes. We then show that higher stability and net charge and changes at the dimer interface were selectively introduced upon separation from old world monkeys and lesser apes (gibbons). Consequently, human, chimpanzee and gorilla SOD1s have a net charge of -6 at physiological pH whereas the closely related gibbons and macaques have -3 . These features consistently point towards selection against the malicious aggregation effects of elevated SOD1 levels in long-living great apes. The findings mirror the impact of human SOD1 mutations that reduce net charge and/or stability and cause ALS, a motor neuron disease characterized by oxidative stress and SOD1 aggregates and triggered by aging. Our study thus marks an example of direct selection for a particular chemical phenotype (high net charge and stability) in a single human protein with possible implications for the evolution of aging.

Key words: primate evolution; evolution of aging; superoxide dismutase 1; oxidative stress; protein aggregation

Introduction

The molecular forces of evolution are responsible for the structure of the proteins[1]: Randomly arising mutations in the DNA are, depending on the population dynamics and the nature of the substitution, either fixated within the population or eventually lost[2][3]. Together with changes in posttranslational properties such as copy number and regulation, which explain some phenotypic differences[4], fixated amino acid substitutions produce the phenotypical differences between species[5][6].

Primate evolution is of importance in the quest towards understanding the origins of the human genome[5][7], and to identify features that make humans unique[8][9]. Non-human primates such as macaques are commonly used to study neurological diseases and identify the relevant differences between primate proteomes[5]. Primate evolution is complex and marked by diversifications due to emergence of generalist strategies relating to changes in habitats and foraging behavior and social innovations[10]. Thus, primate evolution is notable for its variable evolutionary rates caused by geographic diversification, effective population size bottlenecks, and long generation times[11]. Due to their short divergence times, primates show relatively little genetic variation[12]. However, selection for specific gene regulations is well-documented [13] as is ongoing selection on primate genes[14], including human genes[15][16]. In non-human primates, selection has been generally ascribed to genes with reproductive and immunological functions[5]. In humans, most evolutionary changes have been ascribed to the post-translational, epigenetic regime[17].

Among important primate proteins, superoxide dismutase 1 (SOD1) is one of the most abundant proteins of the body[18], with a systemic role in protecting against oxidative stress associated with mitochondrial respiration[19]. SOD1 is a central repressor of cellular respiration by integrating oxygen, superoxide, and glucose signals[20]. It is a dimeric protein with a Cu and a Zn metal ion essential for its catalytic function[19]; its structure is dominated by the β -sheet Greek-Key fold of each dimer, which

evolved from gene duplication events[21]. SOD1 prolongs life by its effects on oxidative stress and metabolism[22–24]. Mutations in SOD1 are risk factors of early-onset familial amyotrophic lateral sclerosis[25–29], which is characterized by age-triggered motor neuron degeneration and mitochondrial dysfunction[30], and SOD1 is a relevant cancer target[31].

Humans display upregulated anti-oxidant defense proteins such as SOD1 upon aging[32][33]. These defenses are tightly linked to the function of SOD1[30][34][35]. We thus speculated that SOD1 is a protein of particular interest to the evolution of great apes, including humans. SOD1 is already known to display a high rate of evolution in primates[36][37]. In 2002, Fukuhara et al. discovered by comparison of 13 SOD1 sequences from primates and rodents that great ape SOD1 evolution is substantially accelerated, and this was interpreted as a sign of positive selection[37]. Recently, the *SOD1* gene was found to be positively selected in humans subject to habitat changes, specifically changes in exposure to high-energy solar radiation and humidity[38]. Thus, SOD1 is one of the very few examples of selection at the protein product (i.e. exon) in humans, yet the molecular mechanism of SOD1 selection has not been identified.

Here we report a detailed study of primate SOD1 evolution using 30 available SOD1 sequences mapped onto the recent consensus mammalian species tree by Meredith et al.[39] and tested for selection and relaxation using a suite of modern codon and branch tests. We identified SOD1 ancestors by computational ancestor reconstruction and annotated key chemical properties of these proteins. We show that SOD1 in great apes has indeed been subject to positive selection as suggested previously from comparison of extant sequences[36][37], and it is the only primate branch showing positive selection. We then show that stability and net charge increased during a remarkably short time frame of ~15 million years due to fast substitutions that also included changes at the dimer interface of SOD1; these properties relate to aggregation and structural integrity of SOD1 as seen in the neurological disease amyotrophic lateral sclerosis (ALS)[28][29][40], which is triggered by human aging and characterized by oxidative

stress and SOD1 aggregation. Our study suggests that SOD1 plays a key role in the evolution of aging and marks a rare example of selection on a molecular phenotype (aggregation propensity) in a human protein.

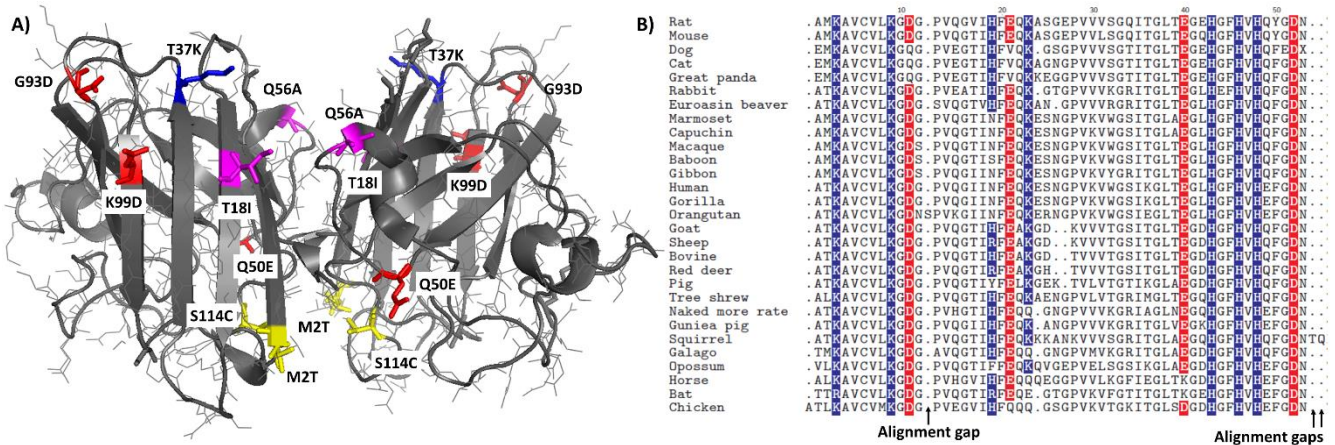


Figure 1. A) Structure of human SOD1 (based on the published structure 2C9V.pdb[41]). The eight sites that experienced substitution within the two branches towards great apes and gibbons and further to great apes are shown in colors and numbered according to the alignment. Red indicates increased net charge (G93D, Q50E, K99D); yellow indicates changes at the dimer interface (M2T, S114C); purple indicates hydrophobic changes at the dimer interface (T18I, Q56A); blue indicates T37K. Of these eight sites, four (T37, Q50, Q56, and S114) underwent positive selection. **B)** First part of the alignment of SOD1 orthologs (full alignment in Supporting Information, Figure S1 and S2). Due to three gaps, numbering increases by 1 from site 13 and by 3 after site 56: For translation to human SOD1, M2T = 2T in human SOD1; T18I = 17I; T37K = 36K; Q50E = 49E; Q56A = 55A; G93D = 90D; K99D = 96D; and S114C = 111C in human SOD1.

Methods

Ancestral reconstruction and phylogenetic analysis

The sequences of 30 mammalian SOD1 proteins available in Uniprot[42] were collected and mapped as extant sequences onto the established mammalian tree with branch-specific orders[39]. There are now SOD1 sequences available that represent all key families, including Hominidae (great apes), Hylobatidae (lesser apes), old world monkeys, and several families of the new world monkeys, as well as sequences from the other suborder of primates Strepsirrhini (galago) and from the closest related order of Scandentia, represented by the tree shrew; this makes the phylogeny well sampled. Such "protein-explicit phylogeny" was previously applied to myoglobin of diving cetaceans[43] and involves mapping specific protein orthologs onto established species trees to deduce the ancestors of the evolved protein, enforced by the species phylogeny. Subsequently, the ancient proteins are studied to identify protein properties that have changed during evolution[43][44]. This approach reveals insight into the molecular causes and types of recent selection pressure on genes.

Nucleotide sequences were aligned using codon-based Clustal Omega[45]. The sequences were unambiguously aligned with only eight generated gaps at different positions (See supplementary information, **Figure S1**). Human and chimpanzee SOD1 sequences are identical, but the sequences of other great apes and primates show variations. Please note that due to the gaps, the alignment numbering differs from the numbering of human SOD1 (**Figure S1**). The best amino acid substitution model, i.e., having lowest Bayesian Information Criteria (BIC) score[46] was found to be that of Whelan and Goldman[47] with among-site-rate-variation[48] modeled by a gamma distribution including the number of invariable sites (WAG+G+I). We then inferred ancestral sequences using the Bayesian model of the MEGA5 software[41]. Computational ancestral protein reconstruction is sensitive to errors when amino acids less likely to be in the ancestral protein still figure prominently in extant sequences. This can affect

the predicted chemical properties and lead to overestimated stability of ancestors because more likely amino acids tend to be more stable[49]. As we use the Bayesian approach this is a minor issue in our work, as also confirmed from the observed opposite trend towards increased stability of extant great ape proteins. For ancestral charges, ancestral state reconstruction is unproblematic because charge of amino acids is not associated strongly with substitution frequencies.

To infer the rate of molecular evolution and evaluate the contribution of selection, we evaluated the ratio of nonsynonymous to synonymous substitution rates, i.e., dN/dS , by different codon models[50]. Codon models fit a set of Markov chains to the observed data (i.e. sequences and the phylogenetic tree) and can accommodate different assumptions regarding rate variation across sequences or branches of the phylogenetic tree. A likelihood function is then evaluated for each model, and one can test different evolutionary hypotheses by comparing the computed likelihood of different nested models[51]. We were particularly interested in investigating i) the evolutionary history of SOD1 of great apes, ii) whether evolution of SOD1 has been driven by distinct types of selection and if so, which branches of the phylogenetic tree and which amino acid residues have been affected, and iii) if these selection types correlate with specific biophysical properties of the enzyme.

We fitted eight different codon models known as M0, FR (free ratio), M1, M2, M3, M7, M8 and M8fix (fixed ratio M8 model)[52]. In brief, M0 models sequence evolution by assuming that all branches have the same rate of evolution, whereas FR fits different rates for different branches. Both M0 and FR are branch models where rate variation at the sequence level is ignored. The remaining models (i.e. M1 to M8fix) are site models that allow the rate of evolution to vary among sites within the sequences. The main differences between these models are whether they assume $dN/dS > 1$ for a fraction of residues and how they model rate variation (e.g. by a beta distribution). While the likelihood of these models were calculated, we compared the likelihoods of M0 vs. FR models to account for rate variation in different

branches of the phylogenetic tree, and M1 vs. M2, M7 vs. M8, and M8 vs. M8fix to establish sites evolving with $dN/dS > 1$.

In addition to the codon models implemented in PAML[53], we also employed the Datamonkey webserver[54] and SLAC (Single Likelihood Ancestor Counting), FEL (Fixed Effects Likelihood), REL (Random Effects Likelihood), MEME (Mixed Effect Model of Evolution)[55] and FUBAR (Fast Unconstrained Bayesian Approximation)[56] models to infer patterns of selection during the evolution of SOD1. The SLAC, FEL and REL methods measure the strength of selection for each codon using different approaches. MEME relaxes the assumption of equal rates among different branches, and FUBAR assumes that each site can belong to a broad distribution of rate-categories. To determine whether elevated rates were due to relaxation of functional constraints or to positive selection, we employed relaxation analysis using the RELAX method[57]. RELAX assigns the significance using a K-parameter to the two cases, relaxation of functional constraints ($K < 1$) and intensification of selection ($K > 1$). Relaxation is characterized by evolutionary rates of different site categories (i.e. both high and low rates) that are skewed towards $dN/dS \sim 1$, whereas intensification of selection is represented by evolutionary rates that are expanded away from $dN/dS \sim 1$.

Computation of protein properties

In order to understand if any specific protein properties were changed significantly during the recent evolution of primates, we studied the two most important chemical properties that affect SOD1 structural integrity, i.e. net charge and thermodynamic stability. These two properties have been shown to play a major role in aggregation of SOD1: Mutations in SOD1 that reduce net charge and stability correlate with more severe clinical outcome of the aging-triggered motor neuron disease ALS, characterized by pathological SOD1 aggregates and molecular oxidative stress[58–61].

To estimate the effect of observed substitutions on thermostability, we used the two methods I-Mutant 2.0[62] and Popmusic 2.1[63], as they have been shown in previous benchmarks to estimate the stability effects of mutations in SOD1 accurately, compared to other methods[64]. All computations were carried out on the SOD1 structure 2C9V, which is of atomic resolution (1.07 Å) required for accurate stability estimates as some stability estimators are sensitive to structure input[65][66].

I-Mutant is a support vector machine method that accounts for T, pH, and the specific local environment of each amino acid within the protein structure, including the solvent accessibility; it was calibrated against experimental data for 2087 single-point mutations[62][67]. Popmusic 2.1 uses energy functions that relate to amino acid volume and solvent accessibility; it was parameterized against a data set of 2648 single-point mutations[63]. All stability changes are reported as a change in free energy of folding caused by substitution i , i.e.

$$\Delta\Delta G_i = \Delta G_i(new) - \Delta G(old) \quad (1)$$

As the template human SOD1 has "old" residues that are sometimes "new" in the phylogeny, changes were assumed to be reversible, as they ideally should be; this approach has previously been shown to be accurate for I-Mutant and Popmusic[43][44], but not generally[68]. In cases where neither the new or old residue was present in the template 2C9V structure but rather a third residue X, we used the fact that free energy is a state function and computed the changes as $\Delta\Delta G$ (new from X) - $\Delta\Delta G$ (old from X). We chose the sign convention that a negative number implies a stabilizing effect of the substitution from old to new. Popmusic 2.1 uses this sign convention whereas I-Mutant 2.0 uses the opposite convention, so these numbers were multiplied by -1 for relative comparison throughout this paper.

To benchmark the ability of these two methods to reproduce general trends in stability changes of mutations occurring in SOD1, we used the normalized $\Delta\Delta G$ values reported by Byström et al.[29], which is the most extensive data set from a single group that we could find. As seen in Supporting

Information, Figure S4, the R^2 values obtained are 0.29 for Popmusic and 0.21 for I-Mutant (numerical data for these benchmarks are given in Supporting Information, Table S4). Whereas the spread is substantial as seen previously[64], the R^2 values show that trend in stability behavior of groups of mutations as observed in the clades studied here are meaningful. Both methods provide the same conclusions of increased stability in great ape SOD1, as shown below. Furthermore, randomly introduced mutations in SOD1 are destabilizing, as expected and generally observed[69], as seen from a computation of all possible 2907 mutations in SOD1 using Popmusic (Supporting Information, Figure S5); only 14% of mutations are stabilizing. This generally means that the observation of stabilizing substitutions is more significant than destabilizing mutations.

SOD1 is negatively charged at physiological pH. The change in net charge due to substitution i (ΔQ_i) was computed simply as:

$$\Delta Q_i = Q_i(new) - Q(old) \quad (2)$$

where $Q_i(new)$ is the charge of the new introduced amino acid residue and $Q(old)$ is the charge of the previous residue in the phylogeny. The charges are -1 for Asp and Glu and $+1$ for Lys and Arg. Because His has a pK_a close to physiological pH it may be both $+1$ and neutral; however, only four histidine-involving substitutions occur in the studied phylogeny, and H20R occurs in the clade leading to bats, whereas L70H and H113N occur in the rodent clade. Only H20N (19N in human SOD1 numbering) occurred in primates upon the split from Strepsirrhini (galagos); this change was assumed to be neutral; had it been positive, the charge would have decreased further towards great apes by 1, strengthening our conclusions. Real charge states of proteins can vary from the theoretically predicted[70–72] and the position of the charge is important; thus even though changes in net charge are well known to affect SOD1 aggregation[70, 73, 74], the positions of the charge changes were also analyzed to get a more complete view as described below.

Results and discussion

General aspects of SOD1 evolution in primates vs. rodents

The orders of rodents and primates separated approximately 75 million years ago[12], and the macaque and human lineages diverged approximately 26 ± 2 million years ago[5]. The speciation of great apes is very dependent on geographical history and has been modeled recently[75]. Rodents constitute the relevant closest related order of primates for comparing evolutionary rates, and accordingly these two orders have previously been compared[76][77]. Primates evolve several times more slowly than rodents[76][77] and humans evolve even more slowly than other primates[5][78][79], a feature that may relate to the longer generation time and reduced mutagenesis rate in great apes due to fewer DNA replications per real time unit[76][80], or alternatively, to lower metabolic rate or smaller historic effective population size; separation of these correlated variables is not straightforward[81][82].

For rodents we find an average of 17.3 non-synonymous substitutions since divergence from the common ancestor of primates and rodents ~75 million years ago, i.e. 1 substitution per 4 million years. For tree shrews and galagos, we identified 13 and 15 substitutions, and for marmoset and capuchin, 17 and 15 substitutions, implying a roughly similar rate. The numbers for baboon (20), macaque (18), and gibbon (18) are also consistent with this approximate rate.

However, for the great apes the picture is different: For orangutan, we observe 27 non-synonymous substitutions in SOD1. For gorilla and human, the numbers are 23 and 22. Comparing the great apes to the remaining primate clade, the two groups of non-synonymous rates (great apes: average 24.0, others: average 16.6) are significantly different (t-test for same mean, assuming different variance,

two-tailed $p = 0.024$, one-tailed = 0.012) (without tree shrew included in primates, the p -values are 0.029 and 0.014, still significant). Thus, among primates, the great apes exhibit significantly more nonsynonymous substitutions.

Since the divide of macaques and humans ~26 million years ago, we observe 13 substitutions, i.e. ~1 fixation per 2 million years, a remarkably higher rate than in rodents in particular considering that rodent rates are generally much higher for other genes [76][80]. Thus, the rate of recent SOD1 evolution of great apes is twice as high as for rodent SOD1, and against the background evolution rates it becomes even more significant. Human SOD1 was previously reported to have evolved quickly, as observed already by Lee et al. and Fukuhara et al. from comparison of sequences of then-known SOD1 orthologs [36][37]. With the substantially expanded sequence data now available and using the more recent consensus tree of mammals[39], we find that this elevated rate is highly statistically significant, and a more detailed study of the evolution of great ape SOD1 thus warrants investigation.

Relaxation of selection in primates and distinct positive selection in great apes

To understand the causes of these elevated rates, we searched for signatures of selection during the evolution of SOD1 in mammals and particularly in primates (see Methods). **Table 1** shows the logarithm of the likelihood function for different codon models. All comparisons are highly significant showing that the evolutionary rate of SOD1 significantly varies among different branches of the mammalian phylogenetic tree, indicating non-neutral evolution. Several site-model comparisons (i.e., M1-M2, M7-M8 and M8-M8fix) show the presence of sites evolving with $dN/dS > 1$ ($p \sim 10^{-12}$). We also looked at models SLAC, FEL, REL, MEME and FUBAR (see Methods and Supporting information, **Table S1**). Similarly, all these models significantly detect non-neutral rate variations across the phylogeny with consistent (three or more) recurrences of sites 18, 27, 33, 35, 43, and 126 using alignment numbering.

Site 27 corresponds to human SOD1 site N26 (without counting the initiator methionine) and is positioned in a very variable segment of 3–5 residues with gaps in the alignment with the triad ESN characteristic of the primates (**Figure S1**). Very interestingly, the highly variable sites 33 and 35 (32W and 34S in human SOD1) are located on opposite sides of a fully conserved G, and these two sites evolved in the branch towards dry-nosed primates (Haplorhini). Site 40 is located alone between two segments that are both highly conserved, and is typically L in primates but D, E or Q in non-primates; it also evolved during the same branch towards Haplorhini.

Table 1. Likelihood of different codon models and significance of comparisons.

| Individual models | | | Comparisons | | |
|-------------------|-----------------|-------------------|-------------|-----------------------------|------------------------|
| Model | np ^a | logL ^b | Comparison | 2 logL1-logL2 ^c | p-value |
| M0 | 59 | -4787.98 | M0-FR | 101.83 | 1.76×10 ⁻⁶ |
| FR | 115 | -4737.07 | | | |
| M1 | 60 | -4501.87 | M1-M2 | 53.90 | 1.99×10 ⁻¹² |
| M2 | 62 | -4474.92 | | | |
| M3 | 63 | -4463.83 | M7-M8 | 52.86 | 3.34×10 ⁻¹² |
| M7 | 60 | -4484.90 | | | |
| M8 | 62 | -4458.47 | M8-M8fix | 41.78 | 1.02×10 ⁻¹⁰ |
| M8fix | 61 | -4479.36 | | | |

a: number of parameters, b: logarithm of likelihood function, c: twice the difference between logarithm of likelihood functions of two nested models.

To check the branches of the phylogenetic tree for positive selection, we used the branch-site test for positive selection in PAML and the BS-REL approach implemented in the Datamonkey webserver (Supporting Information, **Table S2**). These two methods analyze the strength of selection for each codon along each branch of the phylogenetic tree. Therefore, the assumption of equal rate of evolution among branches of the tree (i.e., site-models) or sequences (i.e., branch-models) are removed. Both approaches were corrected for multiple sampling [83]. The branch leading to great apes indicates positive selection ($p \sim 0.023$) with the sites 37, 50, 56, and 114 having high probabilities of selection (see **Figure 1**). These

four sites underwent substitutions towards great apes upon divergence from gibbons: in human SOD1, these selected sites correspond to 37K, 49E, 55A, and 111C. Notably, this is the *only* primate branch showing signatures of positive selection. The posterior probability of positive and purifying selection for each codon site is plotted in **Figure 2A**.

Even though the rate variations are very significant, the underlying evolutionary causes of these higher substitution rates in primates and particularly in great apes could be due to either positive selection in distinct sites of SOD1 *or* due to relaxation of purifying selection already present in the general phylogeny, which would also show up as elevated rates. We thus applied the RELAX approach (see Methods) to determine whether relaxation or intensification of selection pressure occurred. We chose rodents as the reference clade and primates as the test clade. As presented in **Figure 2** and Supplementary Information **Table S3**, this comparison yielded significant differences ($p = 0.013$), showing that the elevated rates in primates relative to what would be expected vs. rodents are partly due to relaxation of some selection pressures. **Figure 2B** shows that the three site-categories ω_1 , ω_2 and ω_3 are relaxed in primates compared to rodents. We do not currently know what these selection pressures are. The relaxation occurs in the early part of the primate phylogeny and could relate to a relaxed need for some features as a trade-off for the increased charge and stability; however we cannot determine this. Thus, we find that relaxation of selection pressures is a significant feature of early primate SOD1 evolution, but that most of the elevated rate of SOD1 evolution in great apes is caused by distinct positive selection as confirmed by various different selection test models (**Table 1**).

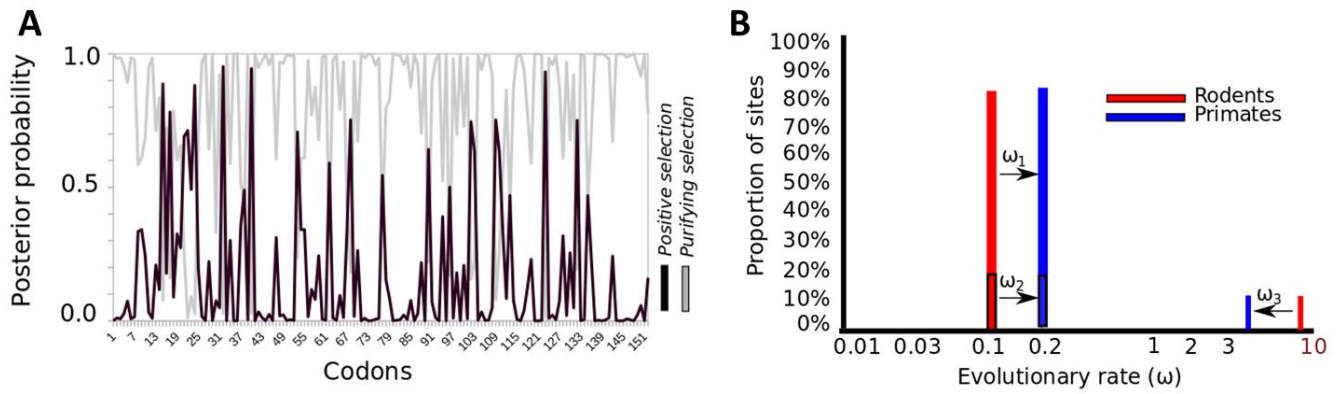


Figure 2. A) Posterior probabilities of positive and purifying selection shown in black and gray respectively versus codon numbers. B) Evolutionary rate transition from rodents to primates for different categories of sites with evolutionary rates ω_1 , ω_2 and ω_3 . (See Table S3 for details).

Increased net charge in SOD1 of great apes

To understand the chemical properties of SOD1 that have been selected during the recent evolution of great ape SOD1, we studied the location of the sites that changed and computed the changes in net charge and stability of the protein, as described in the Methods section. **Figure 1** shows the locations of the eight sites that evolved in the two branches towards great apes and gibbons, and then towards great apes diverging from gibbons. Several observations are notable: i) All these eight sites cluster on the surface of the protein; ii) Four of these sites represent charge changes of SOD1 (T37K, Q50E, G93D, K99D using alignment numbering; corresponding to 36K, 49E, 90D, and 96D in human SOD1) and together increase net charge of human, gorilla, and chimpanzee SOD1 from -3 to -6 ; iii) the remaining four sites (shown in yellow and purple) are all located on the dimer interface; iv) four of the eight sites have signatures from our test models of positive selection (37, 50, 56, 114, corresponding to 36K, 49E, 55A, and 111C in human SOD1) as the only branch in the primate tree. Remarkably, *all* of these chemical property changes relate to the structural integrity of holodimer SOD1 [84][29][73].

somewhat from theoretical estimates[70, 72], but the change of 2 or 3 full charges remains significant as a strong modification of the protein electrostatics. Upon the divergence of the suborders of Strepsirrhini (represented by the galago sequence) and Haplorhini (the remaining sequences), the charge changes were small, -1 towards extant galagos and zero during the Haplorhini branching. The ancestor SOD1 of these suborders (the primate common ancestor SOD1) has a negative charge of -3 , if one discounts any histidines, as discussed in the Methods section. Upon divergence, the new world monkeys (represented by marmoset and capuchins) gained a negative charge to obtain -4 , as the galago SOD1, still not significant. As the old world monkeys (represented by baboon and macaque) diverged from the apes, they experienced K73Q and T138K substitutions which effectively neutralize, so that net charge again remains unchanged; however, in baboons, the K99N (96N in human SOD1) substitution makes SOD1 -4 . Thus, all the monkeys have SOD1 net charges of -3 or -4 . During the branching towards apes, the G93D substitution increased the negative charge of SOD1 of apes to -4 . Importantly, this was again relaxed in gibbons to -3 , whereas in great apes it increased to -6 , due to K99D and Q50E (96 and 49 in human SOD1) compensated by T37K (36K in human SOD1). All these sites are situated on the surface where they would work to increase solvation and repel other SOD1 molecules. One negative charge was relaxed in orangutan, but humans, chimpanzees and gorillas kept the additional negative charges in SOD1. Thus, in humans, gorillas, and chimpanzees, SOD1 has a charge of -6 , the highest in the entire primate clade. If one includes the H20N (19N, human SOD1) substitution as having a charge reduction after divergence from Strepsirrhini, the drift towards higher net charge in Haplorhini and great apes becomes one unit larger, further strengthening this picture.

The observed differences are particularly significant considering that they occurred during a remarkable short evolutionary time. The full charge effect of -3 was introduced into the ancestor of great apes within a matter of 15 million years only, at a rate of non-synonymous substitution of roughly double that of monkeys (**Figure 3**). The increase in net charge from -3 to -6 in this short time span coincides

with detections of positive selection in this branch; the remaining substitutions all occurred at the dimer interface. From **Figure 3** we see that two substitutions increase hydrophobic interaction of the direct dimer interface (T18I, Q56A = 17 and 55 in human SOD1) whereas two others were located at the solvent exposed part of the dimer interface (M2T and S114C = 2 and 111 in human SOD1); both these substitutions are closely interacting in the structure; thus we conclude that increased SOD1 holodimer structural integrity has been specifically selected for in great apes over a relatively short time period.

Changes in thermodynamic stability during primate SOD1 evolution

Because thermodynamic stability is a central feature of SOD1 that preserves its folded structure, and because loss of stability leads to ALS[35][58][61], we also computed the stability changes upon substitution along the phylogenetic tree of primates and rodents as shown in **Figure 4** (see Supporting Information, Table S4 for numerical details). Red implies loss of stability during the transition, whereas blue implies increased stability. For each reported set of numbers, the numbers before and after the slash signifies the *average* $\Delta\Delta G$ computed with I-Mutant 2.0 and Popmusic 2.1 (after correcting for the difference in sign convention of the two methods). Importantly, except for three cases the two methods agree on the sign direction across the phylogeny, despite marked differences in the design of the two methods[63][85]. Furthermore, both methods agree on the direction of change for the whole primate clade except a small difference on the branch towards new world monkeys (marmoset, capuchin) of $+0.1/-0.1$, which is not significant. Both methods also perform well in benchmarks against experimental $\Delta\Delta G$ data for SOD1 [64].

For the two branches where many substitutions occurred, the one leading to Haplorhini and the one leading to great apes, the sign is highly significant and the numeric value of $\Delta\Delta G$ is very likely underestimated; the two branches are both highly stabilizing as computed with both I-Mutant 2.0 and

Popmusic 2.1. The summed of $\Delta\Delta G$ values was computed to be 9.0 and 5.1 kcal/mol for I-Mutant and Popmusic for the first branch towards Haplorhini, and 4.5 and 1.1 kcal/mol for the branch towards great apes. Note that these values are too large and upper bounds to the true stability effect, whereas the average values in **Figure 4** are *lower* bounds, as explained recently [44]. Therefore, only the sign direction has been noted but the numerical change is more substantial (and thus significant) than shown in Figure 4. In contrast, for the branches with 1–2 substitutions the values and even the sign may be uncertain due to the standard error of these methods [64], but these constitute a minor stability effect and the overall drift towards higher thermodynamic stability of SOD1 in great apes is consistent.

To appreciate the observed changes, comparison to another mammalian protein is insightful: Specifically, myoglobin (Mb), a widely studied and highly expressed mammalian protein, has been studied by similar approaches[43], also using the consensus mammalian tree by Meredith et al.[39]. The stability of primate Mb fluctuated but did not increase systematically for apes and primates. In contrast, cetaceans Mb stability *did* increase by 2–3 kcal/mol due to the combination of many substitutions [43]; a finding confirmed by experimental measures of whale vs. human Mb stability[86]. This provides a comparative control where stability of another protein, Mb, is enhanced in cetaceans rather than apes, but stability of SOD1 it is enhanced in great apes. This fits well with the limited role of Mb in primates in terms of oxygen storage, although other functions (such as NO regulation) are notable[87]; in whales, Mb is more important viz. its 10-fold higher concentration and effect on aerobic dive limits[88, 89]. Mirceta et al. showed that net charge of Mb is *not* drifting in terrestrial species including primates and humans but *is* drifting in diving animals such as whales[90]. This provides a control of charge shift, where great apes display no forcing in Mb, whereas they do for SOD1. Intriguingly, the forcing can be related to prevention of protein aggregation in both cases: In whales where Mb concentrations are 10-fold higher than in primates, charge and stability increased for Mb whereas in primates where Mb plays a smaller role but SOD1 is more abundant and plays a large role viz. ALS-related SOD1 mutations, an

opposite trend is observed for the very same properties, plausibly because net charge and stability are *both* important to prevent protein misfolding and aggregation and loss of function of these abundant proteins.

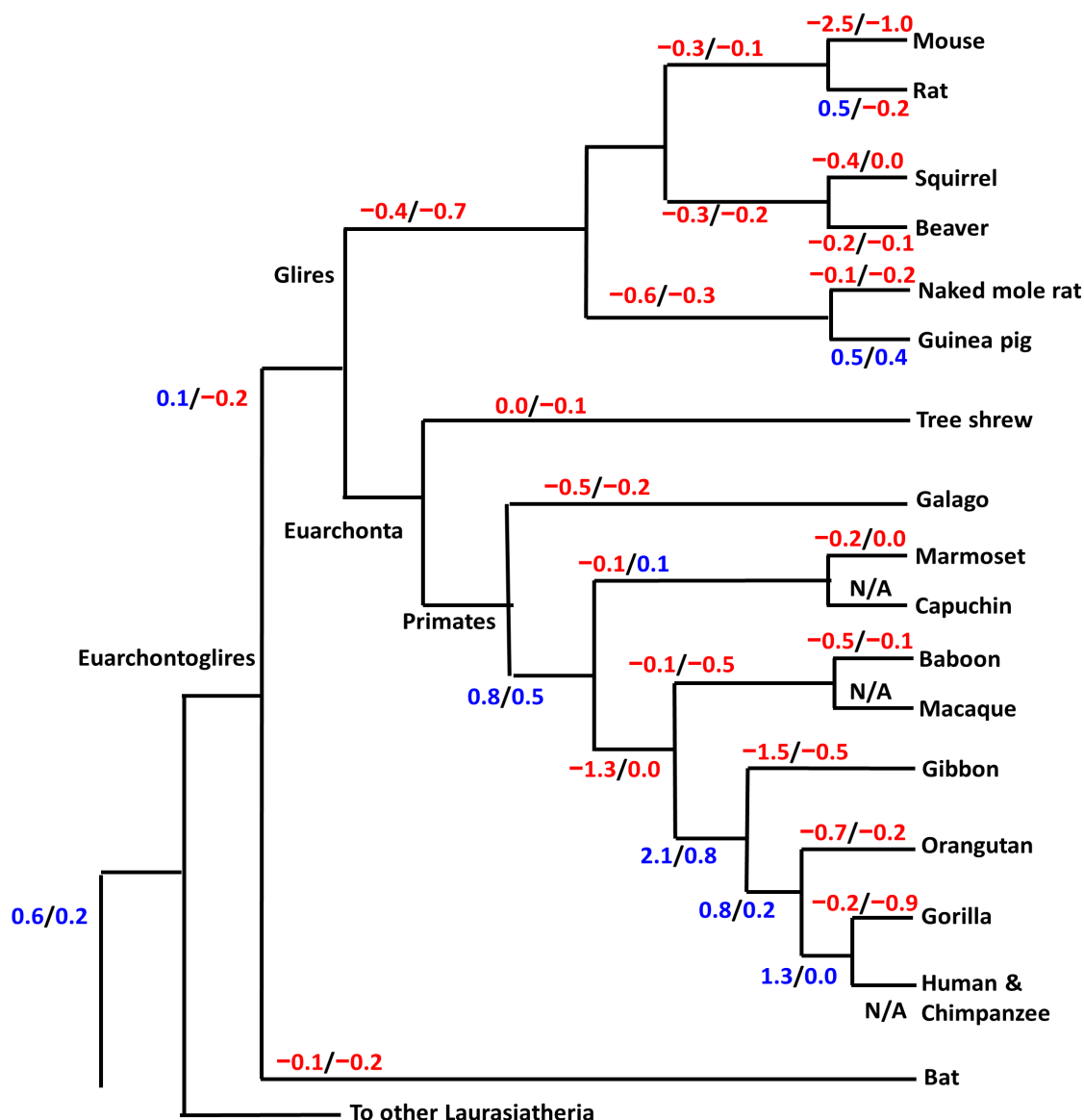


Figure 4. Substitutions occurring in SOD1 of primates, bat, and rodents. As in Figure 3, but in this case only the computed stability changes are shown. Increases in stability are shown in blue whereas reduced stability is shown in red.

Clade-specific differences in effective population size, generation time, and metabolic rate

We have shown that great apes have experienced remarkably elevated rates of SOD1 compared to rodents and other primates. To determine whether the evolutionary rate of a gene in a clade differs from the same gene's rate in another clade, one must account for the general clade-specific differences in evolutionary rates: If the rates were markedly different, but this difference was similar to the general rate difference of the clades, the gene-specific rate difference would most likely be explained by the same general mechanisms that control whole-gene rate differences of the clades.

The clade-specific differences in overall evolutionary rates between rodents, great apes, and other primates relate to the life-history variables of historic effective population sizes N_{eff} , generation times, and metabolic rates of the various species of each clade [81]. Effective population size N_{eff} affects the probability of fixation of arising mutations if evolution is non-neutral, and should then scale linearly with the rate of evolution[91][92]. N_{eff} for mice has been estimated at $\sim 100,000$ [92], about 10-fold that of great apes[93]. At the same time, unfavorable substitutions with a weakly negative selection coefficient are more easily fixated in small populations: Smaller historic N_{eff} would create bottlenecks that affect fixation dynamics. Accordingly, species with small N_{eff} , including some endangered species, have substantially more average genetic variation per individual [14].

N_{eff} values are derived from assumed neutrally evolving DNA segments, which may depend on the generation time and metabolic rate of the organisms; thus the three variables are tightly correlated. The generation times of great apes, in particular humans, are several factors higher than monkeys and macaques[81] and typically ~ 50 times higher than for rodents[81], and their life span is many times longer[94]. Generation time can affect mutation rates since the impact of incident mutations during DNA replication in germ cells is reduced in real time in long-living animals, assuming the same level of DNA repair competency [81][95].

Also, the mass-specific metabolic rate of monkeys is double that of great apes[81]. The specific metabolic rate scales with a power law of the animal mass with exponent between 2/3 and 3/4[96][97], and evolutionary rates and body mass correlate[81]; these effects correlate with generation time[81]. The rate variation can be explained not from N_{eff} but from higher mutation rate in rodents[80], although this has been debated[98]. Multiple regression supported metabolic rate as more explanatory of the evolutionary rate[81], whereas male-female evolutionary rate comparisons in rodents and primates supported a major role of generation time[76]. Regardless of the relative importance of the three features, great apes evolve the most slowly among the studied species in the rodent-primate phylogeny.

Faster SOD1 evolution correlates with slower general evolution

The higher evolutionary rate of SOD1 in great apes than expected viz. the discussion above is remarkable. It is interesting to consider whether the inverse relationship between SOD1 and general rates of evolution has a mechanistic explanation: The ratio of the activity of SOD1 vs. the specific metabolic rate correlates remarkably with life span in a study of 12 primate and two rodent species, suggesting that long-living organisms require more protection against oxidative stress[82]: Monkeys have metabolic rates more than double that of the great apes, and rodents more than five-fold higher; at the same time, specific SOD1 activity is doubly as high in humans as in rodents and intermediate in the monkeys[82]. This suggests that SOD1 plays a role in relation to generation time of great apes. The trend in SOD1 activities is observed both in livers, hearts, and brains[82].

The extended life span of great apes may render protection against oxidative stress more important in real time to maintain the integrity of the genome. SOD1 is well known to a predominant radical scavenger in higher organisms[31][99] and its induction extends lifespan in flies[22][24]. *Together, these anti-aging effects, the higher SOD1 levels of great apes, and the faster evolution rate*

that we observe in SOD1 of great apes, points to a role of SOD1 in an aging program fine-tuned in great apes. Since SOD1 is the central enzyme that converts the superoxide radicals that leak from the oxidative phosphorylation of mitochondria, it is relevant to consider if our observed rate variations in SOD1 evolution can be related to mutagenic rate as done below.

Positive selection for structural integrity in SOD1 of great apes

Our selection analysis showed that SOD1 of great apes has been subject to positive selection. Selection for protein structural integrity acts predominantly on the copy number of misfolded protein molecules [100][101]. The copy number of misfolded and aggregated proteins grows with loss of stability and net charge [102]. Selection pressure to preserve the folded state of abundant proteins conserves highly abundant genes more [101]. SOD1 is not only systemically important in anti-aging functions, it is also one of the most abundant proteins in most cells; thus it makes sense that selection against misfolding acts particularly on SOD1. Furthermore, mass-specific SOD1 levels are 50–90% higher in brains of great apes as in rodents, and are slightly lower in monkeys than in great apes[82].

To explain the clade-specific differences in rodent and great ape evolution, the evolution rate ω in the regime of moderate positive selection can be generally written as[103][104]:

$$\omega = 4 u N_{\text{eff}} s \quad (3)$$

where u is the mutation rate, N_{eff} is the effective population size, and s is the selection coefficient. This equation incorporates the generally much higher evolutionary rates of rodents due to higher N_{eff} or mutation rate. If s is larger for maintaining SOD1 structural integrity in great apes, the rate of evolution of SOD1 should be higher for great apes than for rodents, once corrected for clade-specific evolution rate differences. The relative rates of SOD1 evolution of great apes, $\omega'(\text{SOD1})$, normalized against the

average rates of evolution of the respective genomes, are given by the corresponding ratio of the selection coefficients:

$$\frac{\omega'(SOD1)/\omega(SOD1)}{\omega'(AVG)/\omega(AVG)} = \frac{s'(SOD1)/s(SOD1)}{s'(AVG)/s(AVG)} \quad (4)$$

Here, the unprimed values are for the rodents. Monkeys evolve twice as fast as great apes and rodents typically five times faster than great apes[81]. Since the rates of SOD1 evolution are larger in great apes but within a factor of 2, the selection pressure seems to be at least 2-fold higher in great apes vs. monkeys and 5-fold higher vs. rodents. In the recent theory proposed to explain selection against misfolded proteins[101], the selection coefficient takes the form:

$$s_i = -c_i \Delta U_i = -A_i c_i \left(\exp\left(\frac{\Delta G_i + \Delta \Delta G_i}{RT}\right) - \exp\left(\frac{\Delta G_i}{RT}\right) \right) \quad (5)$$

$$c_i = \frac{N_{aa_i} k_{d_i} (C_{s_i} + C_{d_i})}{\frac{dE_r}{dt}} \quad (6)$$

ΔU_i is the change in copy number of misfolded proteins caused by mutation, A_i is the total abundance of the protein (including folded copies), N_{aa_i} counts the amino acids in the protein, k_{d_i} is the protein degradation rate measured in s^{-1} , C_{s_i} and C_{d_i} are the amino-acid specific costs of synthesizing and degrading SOD1, and $\frac{dE_r}{dt}$ is a scaling number. A typical value of c_i is $\sim 10^{-7}$ [101]. This theory predicts that selection pressure grows monotonically with protein abundance but also with stability and other effects that determine the robustness of the folded pool. Selection for stability is larger if abundance is higher viz. Equation (5), and loss of net charge and dimer integrity would affect the specific cost constant c_i by aggravating turnover costs; indeed, SOD1 mutations that reduce net charge increase aggregation tendency[29, 40, 70, 105]. The costs of Equation (5) can be interpreted in terms of available motor neuron energy [35].

We have shown in Figure 3 and 4 that great ape SOD1 evolved higher net charge and thermostability, which increase structural integrity [106]. Although the physiologically relevant charge state is not directly given by the theoretical charge, the change is highly significant as seen from the many cases where one unit of charge substantially affects aggregation tendency[40, 70, 90]. The physiological relevance of the monomer charge drift from -3 to -5 or -6 is thus quite significant; in the dimer, this change is doubled. Please note that these charge changes occur far from the dimer interface (Figure 1) so their function seems to be mainly to prevent aggregation without affecting dimerization.

As seen from Figure 1, a hydrophobic enhancement at the direct dimer interface is introduced by T18I and Q56A (note that these sites are close in the dimer so they can mutually enhance hydrophobic packing) and polar interactions are introduced at the boundary of the interface by M2T (Figure 1); both are consistent with increased holodimer structural integrity; the enhanced hydrophobic interactions favors association of monomers whereas the polar interactions plausibly prevent alternative interactions of the dimer, although this remains to be investigated. In conclusion, a selective advantage of substitutions towards great apes that increase net charge or stability, according to the findings in **Figure 3** and **Figure 4**, is thus plausible, and this advantage seems related to aging.

The relationship to ALS: Evolution of aging and neurological disease

It is notable that both the protein's net charge and its stability that we have found increase during great ape SOD1 evolution correlate significantly with survival times of ALS patients[29, 35, 61]. SOD1 mutations that reduce stability and net charge of the protein to increase protein misfolding[60][107] are known to cause severe familial forms of ALS[35][61]. SOD1 mRNA levels are also elevated in sporadic ALS, so the role of SOD1 in the motor neuron disease is not limited to the familial cases[108]. The SOD1 mutation toxicity is linked to proteasome function[109] and changes metabolism[110].

Increased proteome maintenance energy costs may contribute to neurodegeneration, since neurons are very energy-demanding and particularly sensitive to energy crisis caused by increased protein turnover (the "energy" explanation of protein misfolding diseases)[35]. Thus protein misfolding may not represent a distinct toxic molecular mode of interaction as generally assumed but perhaps rather the systemic loss of available cellular energy caused by increased protein turnover[101]. This would explain metabolic features of neurological diseases[111] such as the persistent hyper-metabolism of ALS patients[112] and the link between proteasome inhibition and toxicity of ALS-related SOD1 mutants[109], and it indicates why protein misfolding diseases are primarily neurological, since neurons need energy to maintain ion gradients for signaling[35].

Interestingly, the relationship between protein misfolding and energy cost of proteostasis is also of evolutionary importance, and the selection pressure to minimize cellular maintenance energy [101] leaves signatures in the genomes in the form of a bias towards synthetically cheaper amino acids[113][114] and explains why abundant proteins are under purifying selection to evolve slowly[115], as they are associated with larger turnover costs.

One of the distinct features of large animals such as great apes is the longer life span, which as a survival strategy trades with fecundity [116]. Primates are notorious by their extended life span to the limit where aging diseases such as neurodegeneration are common perils. The aging human proteome directly testifies to this aging strategy, by its massive upregulation of anti-oxidant and energy-related gene[32]. Our study shows that great apes experienced selection pressure for the systemic anti-aging protein SOD1 to be more structurally robust to prevent the deleterious effects of protein misfolding. This increased selection pressure may relate to the niching towards specific life-history variables of great apes, but this requires more research to be established.

Concluding remarks

There are three main results of our work: 1) First, great ape SOD1s have experienced significantly higher evolutionary rates compared to other clades in the SOD1 phylogeny ($p \sim 0.024\text{--}0.029$). Considering that gene evolution is normally several times higher in rodents than in primates, and higher in monkeys than in great apes, the tendency of SOD1 to evolve faster becomes even more significant. 2) Second, using a suite of codon models, we identify positive (adaptive) selection in the clade leading to great apes; it is the *only* primate clade showing positive selection for SOD1. Thus, SOD1 is a rare example of positive selection on a human protein product, but confirms indications from a previous study[37] using here newer Bayesian models and a well-resolved phylogeny of newer SOD1 sequences applied to the recent consensus mammal tree[39]. 3) Since divergence from old world monkeys, eight substitutions occurred in great ape SOD1 during ~15 million years (Figure 3); they clustered on the surface and four of them increased net charge by 3 (G93D, T37K, Q50E, K99D) so that the SOD1 monomer went from -3 to -6 ; the four other sites were located at the dimer interface. All these changes relate to the structural integrity of holodimer SOD1[29][84][73], and four of these eight sites have signatures of positive selection. The number of recent substitutions in great apes is substantially higher than in monkeys: The closest relatives, lesser apes (gibbons), did not undergo such changes but kept the net charge unchanged (G93D was compensated by S35R); in orangutan, one charge was lost but SOD1 still has a theoretical charge of -5 , two more than gibbons. Thus the increased net charge is unique to great apes among primates and introduced in a very short evolutionary time coinciding with positive testing for selection.

Great apes have evolved a unique strategy of long life-span with emphasis on raising fewer offspring for complex tasks and knowledge transfer. In comparison to other mammals and rodents, this has led to extremely long life span, small litter size, and long reproductive spans[94]. This specialization also manifests in increased body size and corresponding reductions in metabolic rate and effective population

size. These life-history variables affect the amount of mutation damage incurred over generation time of the gametes and pose a new challenge in the form of aging.

We show that the net charge and the thermodynamic stability of SOD1, which both help to prevent aggregation of the protein[29, 60, 73, 105], have increased during evolution of great ape SOD1. These two properties are known to correlate with the survival time in patients carrying SOD1 mutants causing the neurological aging disease ALS[29, 35, 61]. Altogether, our comparative analysis of SOD1 evolution in rodents and great apes identifies a mechanism that directly frames the evolution of extended aging in great apes within a single protein, suggesting that SOD1 plays a central role in the evolution of aging.

Supporting information

The supporting information file contains the alignment of SOD1 sequences used in this work (Figures S1 and S2); the tree used for rate relaxation analysis as made from DataMonkey (Figure S3); codons detected to be under positive selection using various models (Table S1); branches detected to be under positive selection (Table S2); numerical data from relaxation analysis (Table S3); correlation of benchmarked experimental stability data vs. computed stability changes of SOD1 mutants (Figure S4); numerical data used for this correlation (Table S4); distribution of stability effects for all possible mutations in SOD1 as estimated using Popmusic (Figure S5); all inferred substitutions in the phylogeny from ancestral state reconstruction and computed $\Delta\Delta G$ values and solvent exposure for all sites (Table S5).

References

1. Presgraves DC (2010) The molecular evolutionary basis of species formation. *Nat Rev Genet* 11:175–80. doi: 10.1038/nrg2718
2. Hurst LD (2009) Fundamental concepts in genetics: genetics and the understanding of selection. *Nat Rev Genet* 10:83–93. doi: 10.1038/nrg2506
3. Gillespie JH (1991) *The Causes of Molecular Evolution*. Oxford Ser Ecol Evol. doi: 0195092716
4. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
5. Rogers J, Gibbs RA (2014) Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet* 15:347–359.
6. Wolfe KH, Li W-H (2003) Molecular evolution meets the genomics revolution. *Nat Genet* 33 Suppl:255–265. doi: 10.1038/ng1088
7. Enard W, Pääbo S (2004) Comparative primate genomics. *Annu Rev Genomics Hum Genet* 5:351–378.
8. Goodman M, Grossman LI, Wildman DE (2005) Moving primate genomics beyond the chimpanzee genome. *TRENDS Genet* 21:511–517.
9. Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306–324.
10. Jolly CJ (2001) A proper study for mankind: Analogies from the Papionin monkeys and their implications for human evolution. *Am J Phys Anthropol* 116:177–204. doi: 10.1002/ajpa.10021
11. Locke DP, Hillier LW, Warren WC, et al (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469:529–533.
12. Boffelli D, McAuliffe J, Ovcharenko D, et al (2003) Phylogenetic Shadowing of Primate Sequences to Find Functional Regions of the Human Genome. *Science* (80-) 299:1391–1394. doi: 10.1126/science.1081331
13. Blekhman R, Oshlack A, Chabot AE, et al (2008) Gene Regulation in Primates Evolves under Tissue-Specific Selection Pressures. *PLoS Genet* 4:e1000271.
14. Perry GH, Melsted P, Marioni JC, et al (2012) Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Res* 22:602–610. doi: 10.1101/gr.130468.111
15. Nielsen R, Hellmann I, Hubisz M, et al (2007) Recent and ongoing selection in the human genome. *Nat Rev Genet* 8:857–68. doi: 10.1038/nrg2187
16. Sabeti PC, Schaffner SF, Fry B, et al (2006) Positive natural selection in the human lineage. *Science* 312:1614–20. doi: 10.1126/science.1124309
17. O’Bleness M, Searles VB, Varki A, et al (2012) Evolution of genetic and genomic features unique to the human lineage. *Nat Rev Genet* 13:853–866. doi: 10.1038/nrg3336
18. Pardo C a, Xu Z, Borchelt DR, et al (1995) Superoxide dismutase is an abundant component in

cell bodies, dendrites, and axons of motor neurons and in a subset of other neurons. *Proc Natl Acad Sci U S A* 92:954–8.

19. Perry J, Shin D, Getzoff E, Tainer J (2010) The structural biochemistry of the superoxide dismutases. *Biochim Biophys Acta* 1804:245–262. doi: 10.1016/j.bbapap.2009.11.004
20. Reddi AR, Culotta VC (2013) SOD1 integrates signals from oxygen and glucose to repress respiration. *Cell* 152:224–235. doi: 10.1016/j.cell.2012.11.046
21. Getzoff ED, Tainer JA, Stempien MM, et al (1989) Evolution of CuZn superoxide dismutase and the Greek key β -barrel structural motif. *Proteins Struct Funct Bioinforma* 5:322–336.
22. Landis GN, Tower J (2005) Superoxide dismutase evolution and life span regulation. *Mech Ageing Dev* 126:365–379. doi: 10.1016/j.mad.2004.08.012
23. Sun J, Tower J (1999) FLP recombinase-mediated induction of Cu/Zn-superoxide dismutase transgene expression can extend the life span of adult *Drosophila melanogaster* flies. *Mol Cell Biol* 19:216–228.
24. Melov S, Ravenscroft J, Malik S, et al (2000) Extension of Life-Span with Superoxide Dismutase/Catalase Mimetics. *Science* (80-) 289:1567–1569. doi: 10.1126/science.289.5484.1567
25. Gonzalez de Aguilar J-L, Echaniz-Laguna A, Fergani A, et al (2007) Amyotrophic lateral sclerosis: all roads lead to Rome. *J Neurochem* 101:1153–1160. doi: 10.1111/j.1471-4159.2006.04408.x
26. Rosen DR, Siddique T, Patterson D, et al (1993) Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362:59–62. doi: 10.1038/362059a0
27. Renton AE, Chiò A, Traynor BJ (2014) State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 17:17–23. doi: 10.1038/nn.3584
28. Nordlund A, Oliveberg M (2006) Folding of Cu/Zn superoxide dismutase suggests structural hotspots for gain of neurotoxic function in ALS: parallels to precursors in amyloid disease. *Proc Natl ...* 103:10218–10223.
29. Byström R, Andersen PM, Gröbner G, Oliveberg M (2010) SOD1 mutations targeting surface hydrogen bonds promote amyotrophic lateral sclerosis without reducing apo-state stability. *J Biol Chem* 285:19544–52. doi: 10.1074/jbc.M109.086074
30. Shi P, Gal J, Kwinter DM, et al (2010) Mitochondrial dysfunction in amyotrophic lateral sclerosis. *Biochim Biophys Acta* 1802:45–51. doi: 10.1016/j.bbadis.2009.08.012
31. Huang P, Feng L, Oldham EA, et al (2000) Superoxide dismutase as a target for the selective killing of cancer cells. *Nature* 407:390–395.
32. Lu T, Pan Y, Kao S-Y, et al (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* 429:883–891. doi: 10.1038/nature02661
33. Martin GM, Austad SN, Johnson TE (1996) Genetic analysis of ageing: role of oxidative damage and environmental stresses. *Nat Genet* 13:25–34.
34. Carrì M, Cozzolino M (2011) SOD1 and mitochondria in ALS: a dangerous liaison. *J Bioenerg Biomembr* 43:593–599. doi: 10.1007/s10863-011-9394-z

35. Kepp KP (2015) Genotype-Property Patient-Phenotype Relations Suggest that Proteome Exhaustion Can Cause Amyotrophic Lateral Sclerosis. *PLoS One* 10:e0118649. doi: 10.1371/journal.pone.0118649
36. Lee YM, Friedman DJ, Ayala FJ (1985) Superoxide dismutase: an evolutionary puzzle. *Proc Natl Acad Sci* 82:824–828.
37. Fukuhara R, Tezuka T, Kageyama T (2002) Structure, molecular evolution, and gene expression of primate superoxide dismutases. *Gene* 296:99–109.
38. Hancock AM, Witonsky DB, Gordon AS, et al (2008) Adaptations to Climate in Candidate Genes for Common Metabolic Disorders. *PLoS Genet* 4:e32.
39. Meredith RW, Janečka JE, Gatesy J, et al (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* (80-) 334:521–524.
40. Shaw BF, Moustakas DT, Whitelegge JP, Faull KF (2010) Taking Charge of Proteins: From Neurodegeneration to Industrial Biotechnology. In: *Biology AMBT-A in PC and S* (ed) Adv. Protein Chem. Struct. Biol. Academic Press, pp 127–164
41. Strange RW, Antonyuk S V, Hough MA, et al (2006) Variable metallation of human superoxide dismutase: atomic resolution crystal structures of Cu-Zn, Zn-Zn and as-isolated wild-type enzymes. *J Mol Biol* 356:1152–62. doi: 10.1016/j.jmb.2005.11.081
42. Consortium TU (2008) The Universal Protein Resource (UniProt). *Nucleic Acids Res* 36:D190–D195. doi: 10.1093/nar/gkm895
43. Dasmeh P, Serohijos AWR, Kepp KP, Shakhnovich EI (2013) Positively Selected Sites in Cetacean Myoglobins Contribute to Protein Stability. *PLoS Comput Biol* 9:e1002929.
44. Holm J, Dasmeh P, Kepp KP (2016) Tracking evolution of myoglobin stability in cetaceans using experimentally calibrated computational methods that account for generic protein relaxation. *Biochim Biophys Acta - Proteins Proteomics* 1864:825–834. doi: <http://dx.doi.org/10.1016/j.bbapap.2016.04.004>
45. Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
46. Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6:461–464.
47. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. doi: 10.1093/oxfordjournals.molbev.a003851
48. Yang Z (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 11:367–372.
49. Williams PD, Pollock DD, Blackburne BP, Goldstein RA (2006) Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol* 2:e69. doi: 10.1371/journal.pcbi.0020069
50. Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917.
51. Neyman J, Pearson ES (1992) On the problem of the most efficient tests of statistical hypotheses. In: *Break. Stat.* Springer, pp 73–108

52. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
53. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–91. doi: 10.1093/molbev/msm088
54. Pond SLK, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.
55. Murrell B, Wertheim JO, Moola S, et al (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764.
56. Murrell B, Moola S, Mabona A, et al (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* mst030.
57. Wertheim JO, Murrell B, Smith MD, et al (2014) RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Mol Biol Evol* 32:1–13. doi: 10.1093/molbev/msu400
58. Vassall K a, Stubbs HR, Primmer H a, et al (2011) Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS. *Proc Natl Acad Sci U S A* 108:2210–5. doi: 10.1073/pnas.0913021108
59. Furukawa Y, O'Halloran T V (2005) Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo- and reduced form of SOD1, leading to unfolding and oxidative aggregation. *J Biol Chem* 280:17266–74. doi: 10.1074/jbc.M500482200
60. Lindberg MJ, Byström R, Boknäs N, et al (2005) Systematically perturbed folding patterns of amyotrophic lateral sclerosis (ALS)-associated SOD1 mutants. *Proc Natl Acad Sci U S A* 102:9754–9. doi: 10.1073/pnas.0501957102
61. Wang Q, Johnson JL, Agar NYR, Agar JN (2008) Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol* 6:e170. doi: 10.1371/journal.pbio.0060170
62. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* 33:W306–W310.
63. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M (2011) PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics* 12:151. doi: 10.1186/1471-2105-12-151
64. Kepp KP (2014) Computing Stability Effects of Mutations in Human Superoxide Dismutase 1. *J Phys Chem B* 118:1799–1812.
65. Christensen NJ, Kepp KP (2012) Accurate stabilities of laccase mutants predicted with a modified FoldX protocol. *J Chem Inf Model* 52:3028–3042. doi: 10.1021/ci300398z
66. Kepp KP (2015) Towards a “Golden Standard” for computing globin stability: Stability and structure sensitivity of myoglobin mutants. *Biochim Biophys Acta - Proteins Proteomics* 1854:1239–1248. doi: 10.1016/j.bbapap.2015.06.002
67. Capriotti E, Fariselli P, Casadio R (2004) A neural-network-based method for predicting protein stability changes upon single point mutations. *Bioinformatics* 20 Suppl 1:i63–i68. doi: 10.1093/bioinformatics/bth928
68. Thiltgen G, Goldstein R a (2012) Assessing predictors of changes in protein stability upon

- mutation using self-consistency. *PLoS One* 7:e46084. doi: 10.1371/journal.pone.0046084
69. Tokuriki N, Stricher F, Schymkowitz J, et al (2007) The stability effects of protein mutations appear to be universally distributed. *J Mol Biol*. doi: 10.1016/j.jmb.2007.03.069
 70. Shi Y, Mowery R a, Shaw BF (2013) Effect of metal loading and subcellular pH on net charge of superoxide dismutase-1. *J Mol Biol* 425:4388–4404. doi: 10.1016/j.jmb.2013.07.018
 71. Gao J, Mammen M, Whitesides GM (1996) Evaluating electrostatic contributions to binding with the use of protein charge ladders. *Science* (80-) 272:535.
 72. Shi Y, Abdolvahabi A, Shaw BF (2014) Protein charge ladders reveal that the net charge of ALS-linked superoxide dismutase can be different in sign and magnitude from predicted values. *Protein Sci* 23:1417–1433.
 73. Shaw BF, Valentine JS (2007) How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends Biochem Sci* 32:78–85. doi: 10.1016/j.tibs.2006.12.005
 74. Danielsson J, Mu X, Lang L, et al (2015) Thermodynamics of protein destabilization in live cells. *Proc Natl Acad Sci U S A* 112:12402–12407. doi: 10.1073/pnas.1511308112
 75. Mailund T, Halager AE, Westergaard M, et al (2012) A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet* 8:e1003125. doi: 10.1371/journal.pgen.1003125
 76. Li WH, Ellsworth DL, Krushkal J, et al (1996) Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol Phylogenet Evol* 5:182–187. doi: 10.1006/mpev.1996.0012
 77. Britten RJ (1986) Rates of DNA sequence evolution differ between taxonomic groups. *Science* (80-) 231:1393–1398.
 78. Seino S, Bell GI, Li WH (1992) Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys. *Mol Biol Evol* 9:193–203.
 79. Yi S, Ellsworth DL, Li W-H (2002) Slow molecular clocks in Old World monkeys, apes, and humans. *Mol Biol Evol* 19:2191–2198.
 80. Weinreich DM (2001) The rates of molecular evolution in rodent and primate mitochondrial DNA. *J Mol Evol* 52:40–50.
 81. Martin AP, Palumbi SR (1993) Body size, metabolic rate, generation time, and the molecular clock. *Proc Natl Acad Sci U S A* 90:4087–4091.
 82. Tolmasoff JM, Ono T, Cutler RG (1980) Superoxide dismutase: correlation with life-span and specific metabolic rate in primate species. *Proc Natl Acad Sci* 77:2777–2781.
 83. Anisimova M, Yang Z (2007) Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites. *Mol Biol Evol* 24:1219–1228.
 84. Ding F, Dokholyan N V (2008) Dynamical roles of metal ions and the disulfide bond in Cu, Zn superoxide dismutase folding and aggregation. *Proc Natl Acad Sci U S A* 105:19696–701. doi: 10.1073/pnas.0803266105
 85. Khan S, Vihinen M (2010) Performance of protein stability predictors. *Hum Mutat* 31:675–684.

doi: 10.1002/humu.21242

86. Scott EE, Paster E V., Olson JS (2000) The stabilities of mammalian apomyoglobins vary over a 600-fold range and can be enhanced by comparative mutagenesis. *J Biol Chem* 275:27129–27136. doi: 10.1074/jbc.M000452200
87. Hendgen-Cotta UB, Merx MW, Shiva S, et al (2008) Nitrite reductase activity of myoglobin regulates respiration and cellular viability in myocardial ischemia-reperfusion injury. *Proc Natl Acad Sci U S A* 105:10256–10261. doi: 10.1073/pnas.0801336105
88. Davis RW, Polasek L, Watson R, et al (2004) The diving paradox: new insights into the role of the dive response in air-breathing vertebrates. *Comp Biochem Physiol Part A Mol Integr Physiol* 138:263–268. doi: 10.1016/j.cbpb.2004.05.003
89. Dasmeh P, Davis RW, Kepp KP (2013) Aerobic dive limits of seals with mutant myoglobin using combined thermochemical and physiological data. *Comp Biochem Physiol Part A Mol Integr Physiol* 164:119–128. doi: 10.1016/j.cbpa.2012.10.010
90. Mirceta S, Signore A V., Burns JM, et al (2013) Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science* 340:1234192. doi: 10.1126/science.1234192
91. Kimura M (1962) On the probability of fixation of mutant genes in a population. *Genetics* 47:713.
92. Phifer-Rixey M, Bonhomme F, Boursot P, et al (2012) Adaptive evolution and effective population size in wild house mice. *Mol Biol Evol* 29:2949–2955. doi: 10.1093/molbev/mss105
93. Piganeau G, Eyre-Walker A (2009) Evidence for Variation in the Effective Population Size of Animal Mitochondrial DNA. *PLoS One* 4:e4396.
94. Demetrius L (2006) Aging in mouse and human systems: a comparative study. *Ann N Y Acad Sci* 1067:66–82. doi: 10.1196/annals.1354.010
95. Perez SI, Tejedor MF, Novo NM, Aristide L (2013) Divergence Times and the Evolutionary Radiation of New World Monkeys (Platyrrhini, Primates): An Analysis of Fossil and Molecular Data. *PLoS One* 8:e68029.
96. White CR, Seymour RS (2003) Mammalian basal metabolic rate is proportional to body mass^{2/3}. *Proc Natl Acad Sci* 100:4046–4049. doi: 10.1073/pnas.0436428100
97. Marquet PA (2002) Of predators, prey, and power laws. *Science* (80-) 295:2229–2230.
98. Kumar S, Subramanian S (2002) Mutation rates in mammalian genomes. *Proc Natl Acad Sci* 99:803–808.
99. Scandalios JG (1993) Oxygen Stress and Superoxide Dismutases. *Plant Physiol* 101:7–12.
100. Drummond DA, Wilke CO (2008) Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–352. doi: 10.1016/j.cell.2008.05.042
101. Kepp KP, Dasmeh P (2014) A model of proteostatic energy cost and its use in analysis of proteome trends and sequence evolution. *PLoS One* 9:e90504. doi: 10.1371/journal.pone.0090504
102. Serohijos AWR, Rimas Z, Shakhnovich EI (2012) Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep* 2:249–256. doi: 10.1016/j.celrep.2012.06.022

103. Lanfear R, Kokko H, Eyre-Walker A (2014) Population size and the rate of evolution. *Trends Ecol Evol* 29:33–41. doi: <http://dx.doi.org/10.1016/j.tree.2013.09.009>
104. Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Math Proc Cambridge Philos Soc* 23:838–844.
105. Prudencio M, Hart PJ, Borchelt DR, Andersen PM (2009) Variation in aggregation propensities among ALS-associated variants of SOD1: correlation to human disease. *Hum Mol Genet* 18:3217–26. doi: [10.1093/hmg/ddp260](https://doi.org/10.1093/hmg/ddp260)
106. Dobson CM (2003) Protein folding and misfolding. *Nature* 426:884–90. doi: [10.1038/nature02261](https://doi.org/10.1038/nature02261)
107. Münch C, Bertolotti A (2010) Exposure of hydrophobic surfaces initiates aggregation of diverse ALS-causing superoxide dismutase-1 mutants. *J Mol Biol* 399:512–25. doi: [10.1016/j.jmb.2010.04.019](https://doi.org/10.1016/j.jmb.2010.04.019)
108. Gagliardi S, Cova E, Davin A, et al (2010) SOD1 mRNA expression in sporadic amyotrophic lateral sclerosis. *Neurobiol Dis* 39:198–203. doi: [10.1016/j.nbd.2010.04.008](https://doi.org/10.1016/j.nbd.2010.04.008)
109. Kitamura A, Inada N, Kubota H, et al (2014) Dysregulation of the proteasome increases the toxicity of ALS-linked mutant SOD1. *Genes Cells* 19:209–24. doi: [10.1111/gtc.12125](https://doi.org/10.1111/gtc.12125)
110. Allen SP, Rajan S, Duffy L, et al (2014) Superoxide dismutase 1 mutation in a cellular model of amyotrophic lateral sclerosis shifts energy generation from oxidative phosphorylation to glycolysis. *Neurobiol Aging* 35:1499–509. doi: [10.1016/j.neurobiolaging.2013.11.025](https://doi.org/10.1016/j.neurobiolaging.2013.11.025)
111. Richardson K, Allen SP, Mortiboys H, et al (2013) The effect of SOD1 mutation on cellular bioenergetic profile and viability in response to oxidative stress and influence of mutation-type. *PLoS One* 8:e68256. doi: [10.1371/journal.pone.0068256](https://doi.org/10.1371/journal.pone.0068256)
112. Bouteloup C, Desport J-C, Clavelou P, et al (2009) Hypermetabolism in ALS patients: an early and persistent phenomenon. *J Neurol* 256:1236–42. doi: [10.1007/s00415-009-5100-z](https://doi.org/10.1007/s00415-009-5100-z)
113. Wagner A (2005) Energy constraints on the evolution of gene expression. *Mol Biol Evol* 22:1365–74. doi: [10.1093/molbev/msi126](https://doi.org/10.1093/molbev/msi126)
114. Heizer EM, Raiford DW, Raymer ML, et al (2006) Amino acid cost and codon-usage biases in 6 prokaryotic genomes: A whole-genome analysis. *Mol Biol Evol* 23:1670–1680. doi: [10.1093/molbev/msl029](https://doi.org/10.1093/molbev/msl029)
115. Drummond DA, Bloom JD, Adami C, et al (2005) Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* 102:14338–14343. doi: [10.1073/pnas.0504070102](https://doi.org/10.1073/pnas.0504070102)
116. Kirkwood TB, Rose MR (1991) Evolution of senescence: late survival sacrificed for reproduction. *Philos Trans R Soc Lond B Biol Sci* 332:15–24. doi: [10.1098/rstb.1991.0028](https://doi.org/10.1098/rstb.1991.0028)